Data Models for Interoperability

Qilun Liu[1,2] L.Z. Zhou[1] Jun Chen[2]

[1]*Department of Computer Science and Technology, Tsinghua University*
Beijing, China, 100084
Tel:86-01-62789150
E-mail: liuql@student.cs.tsinghua.edu.cn

[2]*National Geomatics Center of China*
No.1 Zizhuyuan, Baishengcun, Beijing, China, 100044
Tel:86-01-68424120, Fax: 81-01-68424101

ABSTRACT With the development of the computer science, a need for uniform access to and integration of data stored in specialized, non-standard, heterogeneous repositories such as GIS storage servers has become apparent lately. Much effort has been done by some organizations in this area. Interoperability is a solution to this issue. Also, interoperability is a technical aspect of the new concept “Digital Earth”. In this article, we consider different models about data transfer to implement interoperability. Hierarchical model is provided and looked through deeply. Hierarchical model is a synchronize model of the traditional network model and star model. It has inherited the advantages of the both models.
KEY WORDS GIS, Interoperability, data model, the digital earth

1. Background

On January 31, 1998, Vice President of the United States, Mr. Al Gore gave a speech titled “the Digital Earth: Understanding our planet in the 21st Century”. It was the first time to bring about the conception of “the digital earth”.

On June 1,1998, Present Jiang Zemin has argued about Digital Earth on the 9th meeting in the conference of the academy of sciences. He said, “The concept of ‘Information Highway’ were brought about in the past few years, and later on the concept of ‘Knowledge Economics’, and now, the concept of ‘Digital Earth’. The whole society is changing so quickly.”

What is the “Digital Earth”? The Digital Earth refers to a complex system which constructs a geographical space based global information model, and it can describe the information of every point been stored and manipulated in the society of the global conception. In this way, the geographical information is organized efficiently, and everyone can access the geodata through an efficient and convenient query method to have a quick, accurate, and comprehensive information of the whole earth. In the concept of the digital earth, the issue concerned with not only the transfer of data, but it refers to the presentation of Terabytes geodata, multi-scale and 3-dimetion of the earth.

2. Introduction

With the basic conception about Digital Earth, we can learn that there may be many topics to discuss in this scientific field. Interoperability Standards is one of those topics.

The issue of interoperability is a both young and old topic. In the field of database system, the concept of interoperability has been existing for a long time. In Geographical Information System (GIS), it is a new issue because of the rapid accumulation of the geodata, and the growing number of people and organizations who are collecting and using geodata.. Great amount of geodata has been produced everyday at an accelerated rate by data producers and application developers since GIS applications are data-greedy.

Because of the high cost to obtain geographic data, it is now necessary and important to share the geodata. It is then the problem of interoperability of different GIS systems. Because of the data model are different in different systems, especially because of the vector data model. It is because of the special aspect of geodata that different system use different data types. For example, ARC/INFO is using the file system and the connection with the relational DBMS to construct the data model of the

Interoperability is the basic and key issue in implementing the digital earth. Nowadays, many international organization has worked on this area. For example, OGC (Open GIS Consortium) has bring about the OGIS, which include uniform open geodata model, geodata manipulation model and information group model.

In this paper we first explain our understanding of interoperability in GIS and try then to compare traditional data models, at last hierarchical data model is brought about and it is discussed in detail.

3. The Concept of Interoperability

There are many reasons to implement GIS interoperability.

1)multi-semantics

Geodata represent every form of data in GIS, it is inevitable to have multi-semantics. To the same geographical unit, the different applications concerned different semantics: some will regard the natural features such as the longitude and latitude, the height above sea level, climate and so on, others will regard the social features such as the population, economic data etc.

2)multi-temporal and multi-spatial features

Because GIS is concerning the global information, so that all the information has both temporal feature and spatial feature. The data may be of the same time but of different location in a wide area, or, the data may be of the same place at different time.

3)Different scales

The different applications will have different requirement to the precision on the geodata. Also, the different measurement devices can provide geodata with different precision. Therefore, To the different applications, the geodata will have different scales.

4)geodata acquirement

The Geodata include data coming from different sources and different measurement devices such as remote sensor, statistical censor, reconnaissance and so on.

5)storage formats

The Geodata will have both attribute data and topological data. And the topological data have 2 types of the representation: vector and raster. These decide that the geodata will have various storage formats. It is very common that some commercial products like ArcInfo or MapInfo have their own file system to store the topological data.

The term “interoperability” has long been used in the database system fields. In GIS, the definition is focused on the data exchange and function sharing. As Markley defined it as “ the capability for two or more computers to transmit data and to carry out processes as expected by the users.” Gardels describes it as “the ability to access and translate data based on a process of discovery and dynamic interpretation when the salient factors cannot be known in advance Generally, we adopt the concept as Markley has said. To realize interoperability, there are 2 ways as following.

- ***The Approaches of Interoperability***

1)Data Centered Approaches

This method is based on the center data access method. In the data-centered approach, explicit geodata description is especially important as well as rules which allow for encoding objects to a corresponding data access. This approach is what will be discussed in this paper.

2)Process Centered Approaches

In a process-centered approach the specification of methods are standardized mainly through software application programming interfaces(APIs). With online-connections and APIs, it is easier to organize updates. On the other hand, direct access to a global network data may have some problems as security, capacity bottlenecks, and legal constraints.

4.Data.Models.for.Geodata.Interoperability

With the knowledge about the interoperability, we also need some in-depth research about interoperability. In next three steps it will be explained. The first is the structure of interoperability; then the levels of interoperability. At each level there will be different models about the data transfer from one format to another.

- ***The Structure of Interoperability***

In the specifications of OGC, the interope-rability is implemented by using C/S (Client/Server) model. Generally speaking, the interoperability includes 3 parts:

1)Interoperable Gis Data

The object of Interoperable GIS data is to automatize the conversion of the data come from different sources. The data coming from different sources, using different data formats, and being in the different databases should be converted automatically. The process of the conversion must be transparent to the users.

2)Interoperable Gis Applications

The Interoperable GIS applications are used to provide some easy tools to integrate GIS with the other applications. There are 2 kinds about how to integrate them. The first one is loose coupling, in this way, GIS and the other applications are connected by transfering data files. The second way is close coupling, in this way, GIS applications are based on the other applications or vice versa. It is more convenient and comprehensive about the integration.

3)Interoperable Gis Computing

The object of Interoperable GIS computing is to give users the freedom to choose GIS platform in

heterogeneous computing environment. In this way, the computation can be distributed and all the advanced functions can be completed by smallest effort to fit all different circumstances.

In general, the structure of interoperable system can be described in the following graph.

- ***The Levels of Interoperability***

The Levels of interoperability can be considered in 3 levels:

1)Metadata

Metadata is the data of data. The collection of the data is under certain surroundings and according to certain rules. Such information is metadata. It gives us the understanding of the data stored in the database or files. Without metadata, the data are meaningless and useless. It is the most important thing in the data presentation and Interoperability.

2)Data

Data are the representation of the world. There are many kind of data structure and data storage methods. To make different systems interoperable, sharing information is the basic block.

3)Semantic

The object of Interoperable GIS data is to automatize the conversion of data. The data coming from different sources, using different data formats, and being in the different databases should be converted automatically. The data in one application may be meaningless in another. Therefore, it is important to realize semantic interoperability.

- ***Different Data Transfer Models***

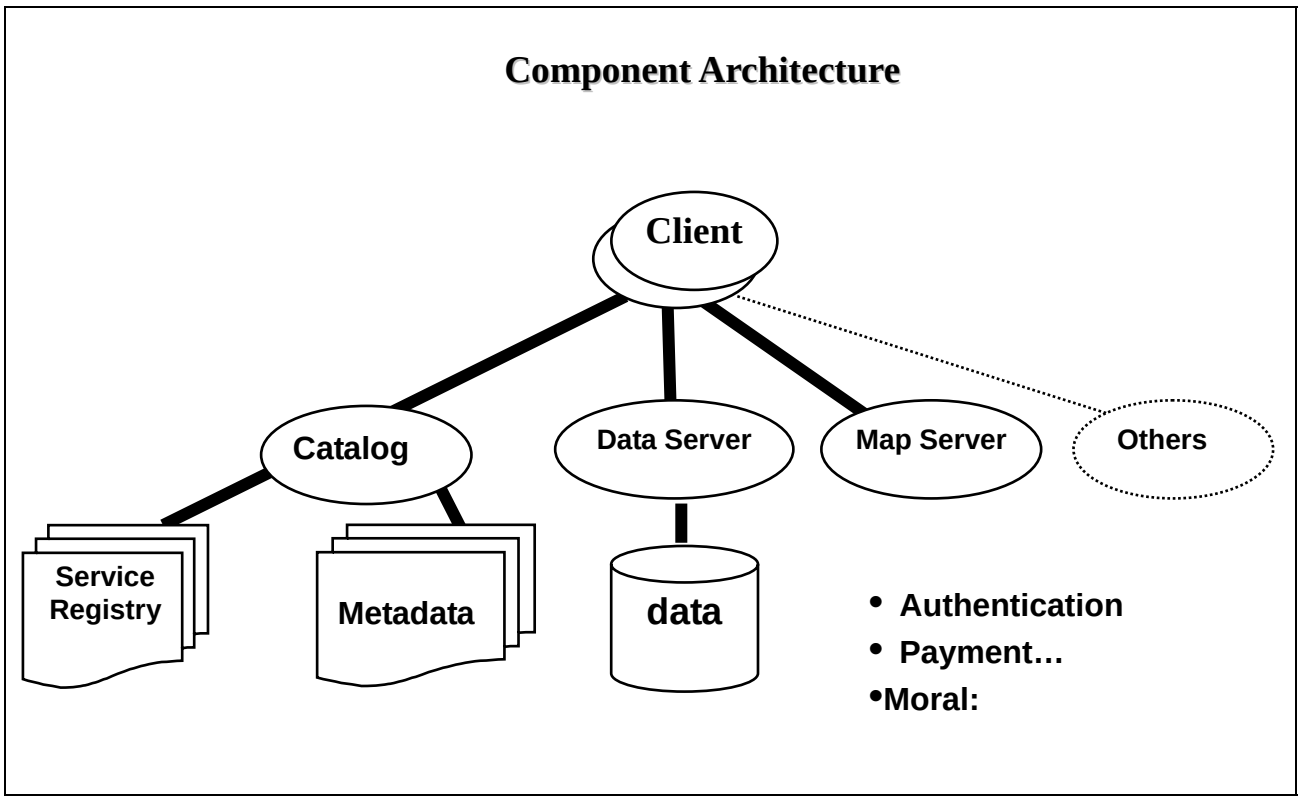
To complete GIS interoperability, one need to implement interoperability in each level. The different level should use different methods. To the first two levels, we can complete interoperability by transfering metadata and data. We will discuss different models to complete the transfering task.

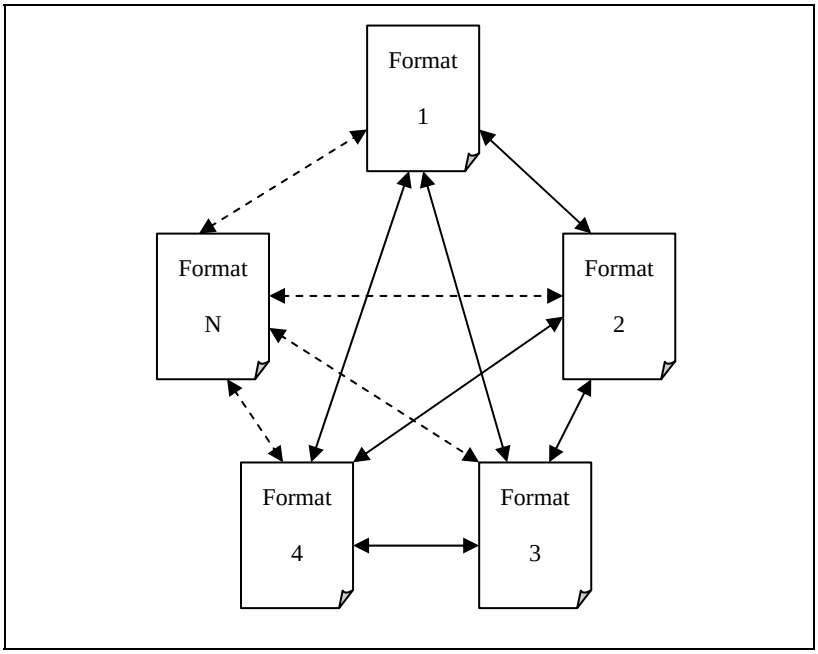
- ***Network model***

The network model is a traditional way to transfer data. It is the simplest way that just writes pairs of data translator between every two formats. In this way, the data can be directly and conveniently translated to another format. The disadvantage of this method is that a translator is needed between each pair of data formats, there must be a number of work to do. Also, it is very difficult to add a new node (data format) into the network. If the N^{th} node is added, then N-1 translators are needed. The extensibility is very bad in this model.

- ***STAR Mode*l**

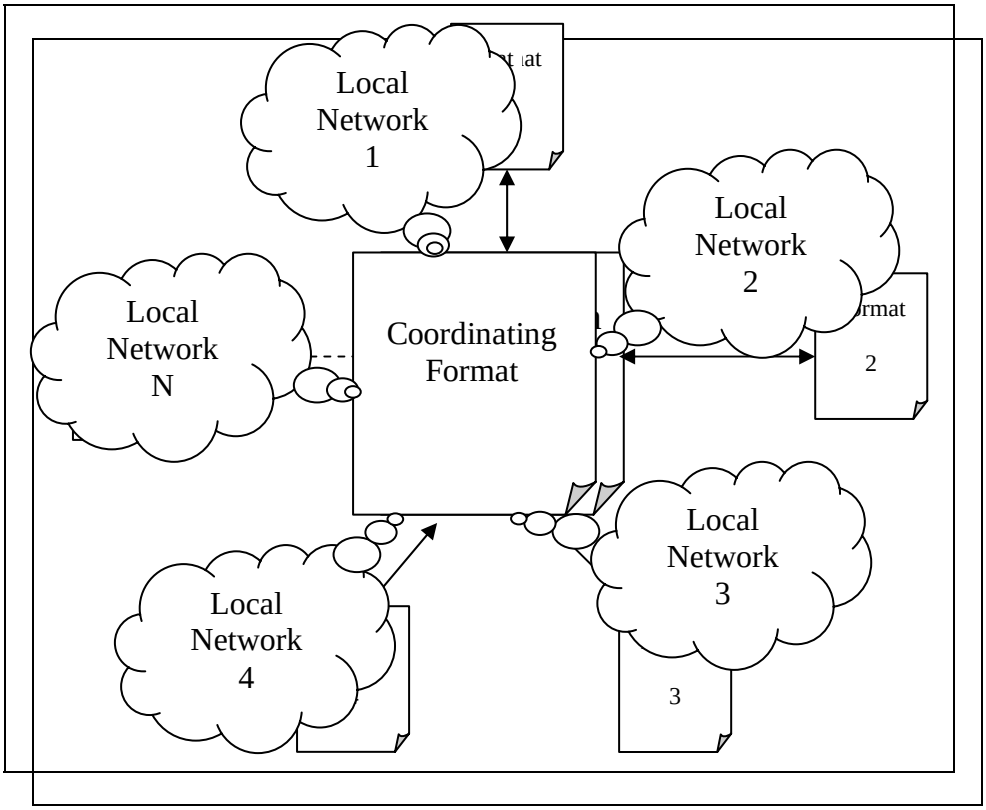
An alternative method to transfer data is called STAR model. As showed in the following graph, there is a Corresponding Format, which every data format can converted to it. With this approach, the





number of data translators can be reduced to a great extent. And the extensibility of this model is very suitable because when a new format is added to the network, only one data translator is needed. These are the advantage of this method. On the other hand, there is few system completed with this method because of the following shortcomings. First, the corresponding format must be very complex to make so many other formats compatible,

so the construction of the corresponding format and the translator to and from the other formats must also be very complicated. Second, the extensibility of the system depends on the completion of the corresponding format. If a new format is not compatible to the corresponding format, then the extensibility cannot be realized, either. Therefore, this way also have some trouble with the data security, because all data are transferred to the corresponding format, so the data can be easily accessed, and the data are not perfectly protected. Also, there will be another problem that the information loss about the two-step conversion from one format to another.



- ***Hierarchical model***

After having discussed the advantages and disadvantages about the above two models, we can find that both model cannot fit our needs very well. For this reason, a new model which is called hierarchical model is submitted here.

First, the construction of the model is like the following. There is a corresponding format similar to the corresponding format in the star model. In this way, different data format can be transferred to another format without too many translators. The local network means the network-like model which mostly consist homogenous data formats. For example, all data formats in “local network 1”are file system based geodata, and all data formats in “local network 2”are relational Database system based geodata. In this way, if two communicating data formats are within the same local network, they may exchange data directly. If they belong to different local network, they need to be first transferred to the corresponding format, then to the object format.

The advantages of this model are very clear. First, the representation within a local network is similar, even though the number of translators may is a way to solve the urgent need for transparent access to heterogeneous data.

In this paper, some concepts and some general knowledge about the digital earth and interoperability are discussed first. After that, some detailed discussion about the data models to implement interoperability are covered. We have analyses advantages and disadvantages about all three models and have come to the conclusion that the model of hierarchical model which is submitted be large, every translator is easy to complete. econd, the loss of information can be reduced since much of the transfer is completed directly, without using a “bridge”—corresponding format. Third, the security can be assured to some extent. Suppose that a local network is within the same department, then the information can be accessed within the local network, while only public data are transferred between different local networks. In this way, the security can be obtained. At last, the model can be extended easily. If the corresponding format is not compatible with a new format, then the original corresponding format can be considered as a local network in a larger size. And with the construction of a new corresponding format, it can be compatible with the new format well. Therefore, a local network can be recursively defined as a set of local networks with a corresponding format. It is clear that this model provides much openness to the structure in interoperability area.

5.Conclusions and Future Research

Digital Earth is a new concept that is brought about recently. Interoperability is one of the technical demands.to.implement.digital.earth. Interoperability

in this paper is better comparing with the other two.

Up to now, no testing prototype has been developed. Much practical work need to be done in the future to have it confirmed if this way can really efficiently and correctly realize the object of interoperability. To complete this task, I'd like to construct an environment with a computer group using different GIS software such as Mapinfo and Arc/info. Multiple OS (operation system) also need to be setup to simulate heterogeneous platform in Internet. Besides, some improvements need to be done to complete the semantic level of interoperability.

Acknowledgement

This research is supported by Natural Science Foun-dation of China (NSFC) under the grant num-ber 69833010

References

OGIS Project Technical Committee, The OpenGIS Guide, Introduction to Interoperable Geoprocessing (1996),Kurt Buehler and Lance Mckee, ed. OGC, Wayland(MA)

Michael F. Goodchild Interoperating GISs 1997.12, Kenn Gardels The Open GIS Approach to Distributed Geodata and Geoprocessing

S.F.Keller, Modeling and sharing geographic data with INTERLIS, Center of Competence INTERLIS, 1998.6